心理学报 2023, Vol. 55, No. 2, 177−191
Acta Psychologica Sinica

https://doi.org/10.3724/SP.J.1041.2023.00177

言语的情绪韵律和情绪语义对听觉去信息掩蔽的作用*

郑 茜[1] 张亭亭[1] 李 量[2] 范 宁[1] 杨志刚[1]

([1]河北大学教育学院，保定 071002) ([2]北京大学心理学系，北京 100871)

摘 要 言语的情绪信息(情绪性韵律和情绪性语义)具有去听觉掩蔽的作用，但其去掩蔽的具体机制还不清楚。本研究通过2个实验，采用主观空间分离范式，通过操纵掩蔽声类型的方式，分别探究言语的情绪韵律和情绪语义去信息掩蔽的机制。结果发现，情绪韵律在知觉信息掩蔽或者在知觉、认知双重信息掩蔽下，均具有去掩蔽的作用。情绪语义在知觉信息掩蔽下不具有去掩蔽的作用，但在知觉、认知双重信息掩蔽下具有去掩蔽的作用。这些结果表明，言语的情绪韵律和情绪语义有着不同的去掩蔽机制。情绪韵律能够优先吸引听者更多的注意，可以克服掩蔽声音在知觉上造成的干扰，但对掩蔽声音在内容上的干扰作用很小。言语的情绪语义能够优先获取听者更多的认知加工资源，具有去认知信息掩蔽的作用，但不具有去知觉信息掩蔽的作用。

关键词 知觉信息掩蔽，认知信息掩蔽，情绪韵律，情绪语义，主观空间分离

分类号 B842

1 引言

在日常生活中，对特定目标说话人声音的加工常会受到来自于其他声源信息的干扰，导致对目标声音加工的困难或失败，这种现象称为听觉掩蔽(auditory masking; Kidd et al., 2005)。已有研究发现，听觉掩蔽主要可以分为能量掩蔽和信息掩蔽两种不同的类型(Arbogast et al., 2002; Brungart, 2001; Freyman et al., 2001)。能量掩蔽主要是目标与掩蔽声音在时间和频率上的重叠所导致的。在特定的时间窗口内，目标声音的特定的频率成分在听觉外周系统(耳蜗和听神经)受到干扰声音的影响，目标声和掩蔽声的成分会共同激活耳蜗基底膜上的相同或邻近区域，导致目标声音在听觉外周的生理表征被破坏，从而导致其后续加工的困难或失败，因此能量掩蔽也被称为外周掩蔽(peripheral masking; Brungart, 2001; Carlile & Corkhill, 2015; Yang et al., 2007)。

然而能量掩蔽并不能解释所有的掩蔽现象。大量研究表明，即使在听觉外周没有产生时频重叠的目标声和掩蔽声进入到听觉中枢系统后，由于掩蔽声和目标声在声学属性和语义内容方面的相似性，掩蔽声与目标声也会在听觉中枢竞争加工资源，造成听者对目标声音加工的困难，这种掩蔽形式被称为信息掩蔽(Kidd & Colburn, 2017; Scott & McGettigan, 2013; Summers & Roberts, 2020; Wilson et al., 2012; Yang et al., 2007)。

然而，一个显见的事实是，听力正常的听者在如“鸡尾酒会”一样的嘈杂环境中，仍具有相对准确的识别目标语音的能力，显示出听觉系统具有很强的抗掩蔽性。这是由于听者能够利用多种线索帮助其对目标声音进行追踪，从而实现对干扰声音的克服(Cherry, 1953)。例如，目标声和掩蔽声在空间位置上的分离(Culling & Mansell, 2013; Glyde et al., 2013)，对目标声音的熟悉感(Huang et al., 2009; Lu et al., 2018)，目标声音的唇读(Wu et al., 2013; Wu

收稿日期: 2022-02-09
* 国家自然科学基金面上项目：鸡尾酒会效应的关键性脑机制(项目批准号：31771252)。
通信作者：杨志刚, E-mail: yangzg.psy@gmail.com

et al., 2017)以及情绪信息等(Dupuis & Pichora-Fuller, 2014; Lu et al., 2018)都被发现具有抗掩蔽的作用。

在日常交流中，说话者的声音通常会带有某种情绪。这些情绪信息既可能体现在声音的韵律上，也可能包含在其语义内容中(Grass et al., 2016; Jaspers-Fayer et al., 2012; Mittermeier et al., 2011)。二者都会对语音的加工过程产生影响。而且与字词的含义相比，声音的音调和节奏更能直接地传递情绪信息(Grass et al., 2016)。

Dupuis 和 Pichora-Fuller (2014)分别考察年轻和年老的成年人在噪音背景下(multi-talker babble)对以 7 种情绪(愤怒，厌恶，恐惧，悲伤，中性，开心，惊喜)发出的单词的再认准确率。结果发现，目标声的情绪韵律确实会影响听者的再认，以恐惧情绪发出的项目的再认准确率最高，其次是以惊喜情绪发出的项目，然而，对以悲伤情绪发出的项目的再认准确率最低，其它 3 种情绪再认的准确率相似。Dupuis 和 Pichora-Fuller 认为听者在噪音中对不同情绪下发出单词的再认行为表现不同，可能是由于情绪声具有明显的可辨别的声学特征。他们根据先前有关情绪的研究，对整个情绪声音的响度范围、持续时间、F0 的平均值和范围进行声学分析。结果发现 7 种情绪的 F0 平均值和范围(最小值和最大值之间的差)差异均显著。以恐惧情绪发出的刺激的 F0 的平均值最高，以惊喜情绪发出的刺激的 F0 的变化范围最大，而以中性情绪发出的刺激的 F0 不论是平均值还是范围均最小。这些分析表明声音韵律的去掩蔽作用可能主要是源于其音高等声学特征，较高的音高会优先在知觉上吸引听者的注意。

然而，另一方面，有研究者认为低层的声学特征并不是导致情绪言语加工优势的唯一原因，情绪性信息本身可能为听者提供了一种趋近动机，使其能够更好地在背景干扰声中检测特定目标声音的存在(Gordon & Hibberts, 2011; Johansson, 1997)。在近期的一项研究中，Lu 等人(2018)在不改变目标声音声学特点的情况下，利用经典条件化范式将原本不具有情绪性的目标声音(中性韵律)与具有显著消极效价的女性的大声尖叫声匹配，从而使目标声音能够诱导出消极情绪。结果发现，经过情绪条件化的声音在信息掩蔽释放量上显著大于与中性声音匹配的声音。这说明即便排除了声学因素的影响后，具有情绪性的声音依然具有显著的去掩蔽作用。

神经影像学的研究发现颞上皮层(superior temporal cortex regions, STR)是对声音韵律进行表征的一个关键区域(Bestelmeyer et al., 2014; Ethofer et al., 2012; Grandjean et al., 2005)。杏仁核是对与危险有关的刺激进行加工的中枢结构。大量的研究表明与中性声音相比，生气声音会引起更大的 STR 和杏仁核的反应(Frühholz & Grandjean, 2013; Grandjean et al., 2005; Sander et al., 2005)。Mothes-Lasch 等人(2016)在以前研究的基础上使用功能性磁共振成像技术(fMRI)探究听觉背景的复杂程度和注意的聚焦点是否能够调节 STR 和杏仁核对于情绪韵律的反应。结果发现，在颞上皮层中部，情绪效应会受到听觉背景复杂程度的影响，在复杂程度低的条件下，对生气情绪韵律发出的目标声音的反应更大；然而，在杏仁核和颞上皮层前部，情绪效应仅受注意的影响，只有将注意集中到韵律声音时，对生气情绪韵律发出的目标声音反应更大。这说明了在不同的脑区，对情绪韵律加工有不同的限制。这一结果也为情绪韵律具有去听觉掩蔽的作用提供了神经机制方面的证据。

另一方面，言语中所包含的语义情绪信息具有加工优势，已经为大量研究所证实(Goh et al., 2016; Iwashiro et al., 2013)。例如，Goh 等人使用词汇判断任务和语义分类任务考察单词语义的情绪效价是否对单词的再认产生影响。在词汇判断任务中，要求听者尽可能快且准确地判断双耳听到的声音是词还是非词；在语义分类任务中，要求听者尽可能快而准确地判断双耳听到的单词在语义上是具体的还是抽象的。结果发现：在两个任务中，与中性单词相比，听者对消极语义单词和积极情绪语义单词的反应都更快。

此外，Iwashiro 等人(2013)使用具有情绪语义的单词诱导情绪。除情绪效价(消极、中性、积极)外，单词的唤醒度、优势度和词频差异均不显著；并且由合成的中性韵律发出单词，不同效价的单词的振幅和音高在平均值、方差和范围上差异均不显著。所有单词的呈现时间和 RMS 相同。使用双耳分听范式，左耳呈现消极、中性、积极单词，右耳呈现中性单词；或者相反。要求听者有选择地注意左耳或者右耳，并且尽可能快地从屏幕上出现的 4 个单词中选择注意耳所听到的单词。结果表明，无论是否将注意集中在消极单词上，听者在有消极单词出现时的反应时间均显著长于两耳均呈现中性单词时的反应时间。这说明不论是在注意还是在非注意条件下，听觉呈现的消极情绪词都会占用听者

更多的加工资源。

那么言语中所包含的语义情绪信息在有干扰的环境中是否也具有加工优势, 可以起到去掩蔽的作用? 如果有的话, 其在机制上与情绪韵律线索的去掩蔽机制是否有所不同?

如前所述, 听觉掩蔽中包含着不同的成分, 其中能量掩蔽不受认知、注意等高级加工的调节, 具有很强的不可逆性(Wu et al., 2007; Yang et al., 2007)。因此, 言语的韵律和语义中所包含的情绪信息的去掩蔽作用很可能是主要作用于信息掩蔽而非能量掩蔽。但另一方面, 信息掩蔽本身也可能有不同的发生机制。虽然很长一段时间内它都被当作一个行李箱词汇, 但近年来已经有越来越多的研究者指出信息掩蔽中可能存在着不同的成分(杨志刚等, 2014; Carlile & Corkhill, 2015; Cooke et al., 2008; Mattys et al., 2009; Watson, 2005)。比如, 杨志刚等(2014)认为信息掩蔽可以分为知觉信息掩蔽(perceptual informational masking)和认知信息掩蔽(cognitive informational masking)两种成分。知觉信息掩蔽是指掩蔽声音与目标声音竞争注意资源所导致的干扰; 认知信息掩蔽是由掩蔽声音与目标声音竞争认知/言语加工资源所导致的。考虑到前面所提出的言语情绪韵律与情绪语义的特性, 是否两种线索可以对不同的信息掩蔽子成分发挥作用? 具体地说, 具有情绪性韵律的声音能够更好地吸引被试的注意力, 使其能够在混合的声音流中被追随, 因而有利于克服知觉掩蔽; 而目标声音中所包含的情绪语义信息则会因其特殊的生态学意义而得到中枢系统的优先激活, 从而会有助于减少对其的认知掩蔽。因此, 本研究拟探究言语的情绪韵律和情绪语义在去知觉信息掩蔽和去认知信息掩蔽中的作用。从实际意义来看, 在现实的交流场景中, 不论是情绪韵律还是情绪语义, 都是在交流中经常会出现的泛语言元素, 了解这元素在交流中的作用和具体机制, 对于提升日常交流的效能, 以至于开发更为有效的人工耳蜗算法都有重要的启示。

已有研究表明, 对于英语和汉语目标语音的再认, 当背景掩蔽声数量为 2 个时, 造成的信息掩蔽量最大(Freyman et al., 2004; Freyman et al., 1999; Rakerd et al., 2006; Wu et al., 2007)。因此, 本研究使用两个掩蔽声来模拟“鸡尾酒会”听觉条件。此外, 时间逆转语句不具有语义可懂度, 仅能够产生知觉信息掩蔽; 而正序的无意义语句具有一定程度的语义可懂度, 既可以造成知觉掩蔽, 也可以造成认知掩蔽。本研究使用这两种语句作为掩蔽材料, 从而可以对知觉掩蔽与认知掩蔽的作用进行分离。主观空间分离范式利用优先效应能够使听者主观感知到目标声和掩蔽声来自不同的位置(虽然实际上目标声和掩蔽声播放位置相同), 其主要释放信息掩蔽, 对能量掩蔽影响不大(Freyman et al., 1999, 2001)。因而, 本研究引用主观空间分离范式探究情绪韵律与情绪语义线索去信息掩蔽的机制。

2 实验 1: 情绪韵律的去掩蔽作用

实验 1 考察情绪韵律的去掩蔽作用。利用知觉空间分离下多个说话人声音掩蔽范式, 分别在实验 1a 和实验 1b 中考察逆序与正序言语掩蔽条件下目标言语的情绪韵律的去掩蔽作用。

2.1 实验 1a: 逆序言语掩蔽条件下情绪韵律的去信息掩蔽作用

2.1.1 被试

研究 1a 的被试为 26 名听力正常的河北大学在校学生, 其中男生 10 人, 女生 16 人。平均年龄为 19.85 ± 2.15 岁。母语为汉语, 右利手。所有被试均通过纯音听力测试(听力计, Conera, GN OTOMETRICS A/S), 具有正常且左右平衡的听力(左右耳相差不高于 15 dB, 在任何一个频率上不高于 20 dB)。实验结束后, 获得一定报酬。

2.1.2 材料和设备

目标刺激为句法正确的汉语无意义句子(Yang et al., 2007), 每个句子都包含 12 个音节。其中包括三个关键成分: 主语、谓语和宾语。这三个成分同样是三个关键词, 每个关键词由两个音节组成, 如“他的**水道**可能**停放**这个**动脉**”。实验前由 35 名非实验被试对词汇(动词: 571、名词: 1091)的效价和唤醒度进行 9 点评分。动词、名词效价的有效评分者分别为 26 个和 29 个。所有的词汇材料均来自中文词汇数据库(CLD) http://www.chineselexicaldatabase.com/ (Sun et al., 2018)。选择效价评分为 4.50~5.50, 唤醒度相对较低的动词(182 个, 评分 ≤ 6.00)和名词(364 个, 评分 ≤ 3.51)作为中性词构成目标语句。之后所有目标语句均由同一个年轻女性(C)以不同情绪韵律(中性、开心)读出并进行录音, 每句时长为 3 s 左右。录制后的句子由 20 个人(女, 10)对声音韵律的效价和唤醒度进行 5 点评分。结果表明, 中性韵律的效价平均分为 1.71 ± 0.23, 唤醒度为 2.20 ± 0.24; 开心韵律的效价平均分为 4.20 ± 0.21, 唤醒度为 4.06 ± 0.20。中性韵律和开心韵律

语句在效价和唤醒度上均差异显著[效价：F (1, 180) = 5793.64, p < 0.001; 唤醒度：F (1, 180) = 3147.26, p < 0.001]。

掩蔽刺激为 2 个说话人说出的时间逆转的无意义语句。逆序言语具有与正常言语相反的时间变化模式，但是能保留人们嗓音的特征，同样有谐波和共振峰结构。将两个女性(A 和 B)读出的句法正确的汉语无意义句子，经 Matlab 程序在时间上逆转后，按照不同的实验条件进行随机等强度叠加，形成 2 个说话人的混合声音。时间逆转言语不具有语义可懂度，被认为仅能够产生知觉信息掩蔽(Yang et al., 2007)。

所有声音信号都用声音编辑软件 Adobe Audition 以 22.05 kHz 的采样率数字化，并形成 16-Bit 的 PCM 波形文件。运用心理学实验软件 Matlab 对实验使用的刺激进行合成。目标声音在单耳耳机中的呈现水平为 56 dB，以目标声的声压级为基准调整掩蔽声的呈现水平以得到规定的信噪比。

实验时刺激由 Matlab 呈现。声音刺激由电脑声卡发出，通过双声道耳机(isk, HP-960B)呈现给被试。目标声和掩蔽声均在左右两个声道中播放，在有主观空间分离的条件下，目标声在右侧声道中起始的时间领先其在左侧声道中的相同复本 3 ms，而掩蔽声在左侧声道中领先其在右侧声道中的相同复本 3 ms，如此使得目标声听起来是在右侧声道中发出，而掩蔽声在左侧声道中发出，两者有主观(知觉的)的空间分离；在没有主观空间分离的条件下，目标和掩蔽声均同时在左右声道发出，没有时间延迟。

2.1.3　设计与程序

实验采用 2×2×4 的被试内设计。3 个被试内变量及水平分别为：(1)主观空间分离(有主观空间分离、无主观空间分离); (2)情绪韵律(中性韵律、开心韵律); (3)信噪比(−8 dB、−4 dB、0 dB、4 dB)。每个被试共接受 16 个实验条件，每个条件下有 11 个 trials。根据主观空间位置关系和情绪韵律条件，将所有 trials 分为 4 个 block, 4 个 block 的顺序在不同的被试间进行完全拉丁方平衡。4 种信噪比水平在每个 block 中随机化。因变量为被试复述关键词的正确率。

被试进入实验室后被要求坐于电脑前，头与电脑屏幕平行，不得随便移动。主试给被试讲解指导语，为确保被试能完全理解指导语并能正确跟踪目标语句，在正式实验之前被试会接受大约 10 分钟的练习。练习中首先在安静条件下呈现目标声音，让被试熟悉目标声音的音色；再给出几个练习试次进行练习。练习阶段使用的语句和正式实验阶段使用的语句均不同。

练习之后进入正式实验。被试按下电脑键盘的空格键开始实验。在每个试次中，掩蔽声音呈现 1 s 之后，目标语句呈现，并和掩蔽声同时结束。被试在每个试次结束后，要尽可能地复述整个目标语句。实验人员坐在被试旁，记录被试重复的关键词，每个关键词两个音节，复述对一个音节记一分。被试完成一个 block 之后，休息 2 分钟，再完成下一个 block。整个实验持续大约 40 分钟。实验结束后主试对被试在每种条件下的识别正确率进行离线计算。

2.1.4　结果与分析

首先，对所有被试在不同条件下识别目标句的正确率进行均值分析，描述统计结果见表 1。

表 1　识别目标句的正确率(M ± SD)

信噪比(dB)	无主观空间分离		有主观空间分离	
	中性韵律	开心韵律	中性韵律	开心韵律
−8	0.29 ± 0.16	0.39 ± 0.14	0.46 ± 0.19	0.50 ± 0.16
−4	0.55 ± 0.18	0.66 ± 0.15	0.71 ± 0.16	0.71 ± 0.19
−0	0.76 ± 0.16	0.80 ± 0.10	0.87 ± 0.08	0.86 ± 0.09
−4	0.93 ± 0.04	0.90 ± 0.07	0.94 ± 0.04	0.94 ± 0.04

对不同实验条件下的识别正确率进行 3 因素的重复测量方差分析(所有 p 值都经过 Greenhouse-Geisser 矫正)，结果表明，主观空间分离主效应显著，F (1, 25) = 102.66, p < 0.001, η_p^2 = 0.80，有主观空间分离条件下的识别正确率(M = 0.75, 95% CI: [0.72, 0.79])高于无主观空间分离条件下的识别正确率(M = 0.66, 95% CI: [0.62, 0.70])。情绪韵律主效应显著，F (1, 25) = 6.02, p = 0.02, η_p^2 = 0.19，开心韵律条件下的识别正确率(M = 0.72, 95% CI: [0.68, 0.75])高于中性韵律条件下的识别正确率(M = 0.69, 95% CI: [0.65, 0.73])。信噪比主效应显著，F (3, 75) = 314.75, p < 0.001, η_p^2 = 0.93，随着信噪比的提高，识别的准确率也提高，ps < 0.001。

主观空间分离、情绪韵律和信噪比三者交互作用显著，F (3, 75) = 3.16, p = 0.041, η_p^2 = 0.11。随后的简单简单效应分析发现，在无主观空间分离且信噪比为−8 dB 或者−4 dB 时，开心韵律条件下的识别正确率显著高于中性韵律条件下的识别正确率(p = 0.001; p = 0.002)。

为了进一步得到在知觉信息掩蔽下，情绪韵

律、主观空间分离线索和二者结合时的掩蔽释放量，使用Origin 8.0软件中的Logistic心理测量函数，对每个被试在不同实验条件下识别目标语句的正确率进行拟合，并计算其在50%识别正确率时的识别阈限(Lu et al., 2018)。

$$y = \frac{1}{1 + e^{-\sigma(x-\mu)}}$$

在该函数中，y 是被试能正确识别关键词的概率，x 是与 y 相对应的信噪比，μ 是指识别目标语句的正确率为 50%处的信噪比，即语音识别阈限，σ 是与 μ 对应的斜率。图 1 显示了所有被试平均识别率的拟合结果。不同主观空间分离和情绪韵律条件下平均识别阈限结果如图 2 所示。

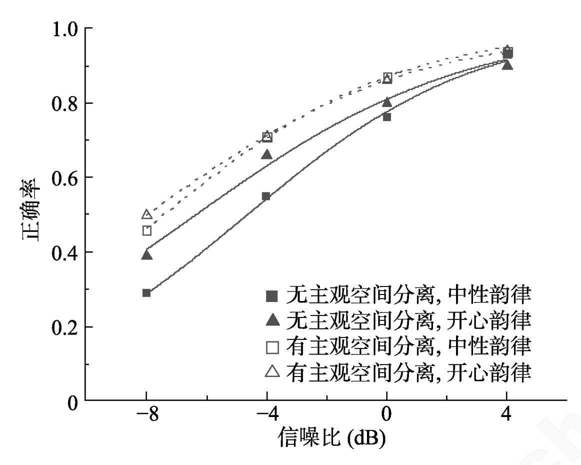


图 1　逆序言语掩蔽下目标关键词的识别率

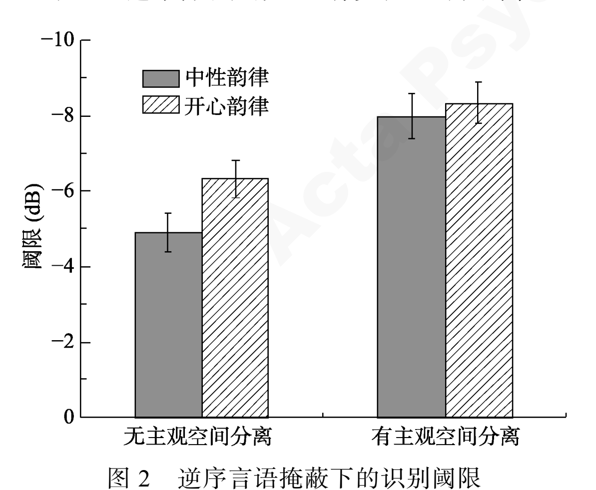


图 2　逆序言语掩蔽下的识别阈限

为了分别考察情绪韵律线索和主观空间分离线索的去掩蔽效应，参照前人研究(Lu et al., 2018)，将在无主观空间分离条件下目标语句以中性韵律发出时的识别阈限减去以开心韵律发出时的识别阈限，得到单独的情绪韵律线索的掩蔽释放量；将目标语句以中性韵律发出时，无主观空间分离条件下的识别阈限减去有主观空间分离条件下的识别阈限，得到单独的主观空间分离线索的掩蔽释放量；将无主观空间分离且目标语句以中性韵律发出时的识别阈限减去有主观空间分离且目标以开心韵律发出时的识别阈限，得到情绪韵律线索和主观空间分离线索结合时的掩蔽释放量。

对所得三种掩蔽释放量进行单因素方差分析，结果表明，听觉线索类型效应显著，F (2, 75) = 6.03, p = 0.004。情绪韵律的掩蔽释放量(M = 1.41, 95% CI: [0.67, 2.15])显著低于主观空间分离(M = 3.06, 95% CI: [2.32, 3.80], p = 0.009)和二者结合的掩蔽释放量(M = 3.42, 95% CI: [2.26, 4.59], p = 0.002)；但主观空间分离的掩蔽释放量和二者结合的效应差异不显著，p = 0.557。结果如图 3 所示。

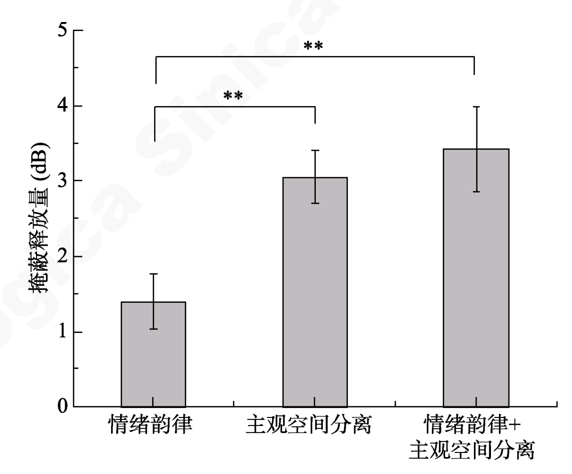


图 3　不同听觉线索的掩蔽释放量

注：*p < 0.05, **p < 0.01, ***p < 0.001，下同。

综上，实验 1a 发现，在逆向言语掩蔽条件下，情绪韵律和主观空间分离均具有去知觉信息掩蔽的作用，但前者的作用要比后者更小。且情绪韵律的作用主要是在没有主观空间分离且信噪比比较低的条件下才显现出来。这可能是由于在知觉条件相对困难的情况下，更需要被试利用任何可得的线索帮助其追踪和识别目标声音。而在相对简单的识别条件下，如有空间分离或信噪比相对较高，主观空间分离和信噪比线索就已经释放了大量的知觉掩蔽，目标语音的情绪韵律作为一种知觉线索的作用就无从发挥了。接下来的实验 1b 将使用句法正确的正序无意义语句作为掩蔽材料，探究情绪韵律在知觉、认知双重信息掩蔽下，是否还具有去掩蔽的作用。

2.2　实验 1b：正序言语掩蔽条件下情绪韵律的去信息掩蔽作用

2.2.1　被试

31 名听力正常的河北大学在校学生参加研究，

其中男生 15 人，女生 16 人。平均年龄为 19.45 ± 1.39 岁。母语为汉语，右利手，并通过纯音听力测试(听力计，Conera, GN OTOMETRICS A/S)。所有被试均具有正常且左右平衡的听力(左右耳相差不高于 15 dB，在任何一个频率上不高于 20 dB)。实验结束后，获得一定报酬。

2.2.2　材料和设备

目标刺激同实验 1a，掩蔽刺激为正序播放的 2 个说话人说出的句法正确的汉语无意义语句。这种声音被认为能够产生知觉、认知双重信息掩蔽(Yang et al., 2007)。实验设备同实验 1a。

2.2.3　设计和程序

均同实验 1a。

2.2.4　实验结果

使用 SPSS 21.0 软件对所有数据进行统计分析。

首先，对所有被试在不同条件下识别目标句的正确率进行均值分析，描述统计结果见表 2。

表 2　识别目标句的正确率($M \pm SD$)

信噪比(dB)	无主观空间分离		有主观空间分离	
	中性韵律	开心韵律	中性韵律	开心韵律
−8	0.07 ± 0.07	0.18 ± 0.11	0.28 ± 0.18	0.37 ± 0.14
−4	0.22 ± 0.14	0.43 ± 0.15	0.58 ± 0.19	0.63 ± 0.12
0	0.55 ± 0.19	0.67 ± 0.09	0.79 ± 0.13	0.81 ± 0.09
4	0.86 ± 0.12	0.85 ± 0.07	0.90 ± 0.07	0.91 ± 0.06

对不同实验条件下的识别正确率进行 3 因素重复测量方差分析(所有 p 值都经过 Greenhouse-Geisser 矫正)。结果表明，主观空间分离主效应显著，F (1, 30) = 357.27, p < 0.001, $\eta_p^2 = 0.92$，有主观空间分离条件下的识别正确率(M = 0.66, 95% CI: [0.63, 0.69])显著高于无主观空间分离条件下的识别正确率(M = 0.48, 95% CI: [0.45, 0.51])。情绪韵律主效应显著，F (1, 30) = 38.15, p < 0.001, $\eta_p^2 = 0.56$，开心韵律条件下的识别正确率(M = 0.61, 95% CI: [0.58, 0.63])高于中性韵律条件下的识别正确率(M = 0.53, 95% CI: [0.49, 0.57])。信噪比主效应显著，F (3, 90) = 764.63, p < 0.001, $\eta_p^2 = 0.96$，随着信噪比的提高，识别准确率也提高，ps < 0.001。

主观空间分离、情绪韵律和信噪比三者交互作用显著，F (3, 90) = 5.21, p = 0.003, $\eta_p^2 = 0.15$。简单简单效应分析发现，在无主观空间分离且信噪比为 −8 dB、−4 dB 或者 0 dB 时，开心韵律条件下的识别正确率显著高于中性韵律条件下的识别正确率，ps < 0.001；在有主观空间分离且信噪比为−8 dB 时，开心韵律条件下的识别正确率显著高于中性韵律条件下的识别正确率，p < 0.001。

同实验 1a，将每个被试在不同实验条件下的识别目标语句的正确率转化为相应的语音识别阈限。图 4 显示了所有被试平均识别率的拟合结果。不同主观空间分离和情绪韵律条件下平均识别阈限结果如图 5 所示。并进而求得在知觉、认知双重信息掩蔽下，情绪韵律线索、主观空间分离线索和二者结合时的掩蔽释放量。

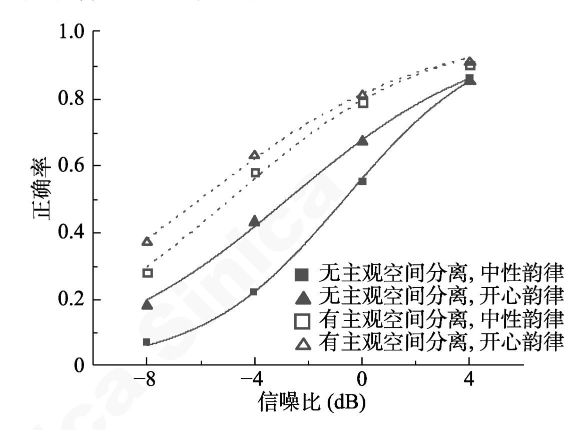


图 4　正序言语掩蔽下目标关键词识别正确率

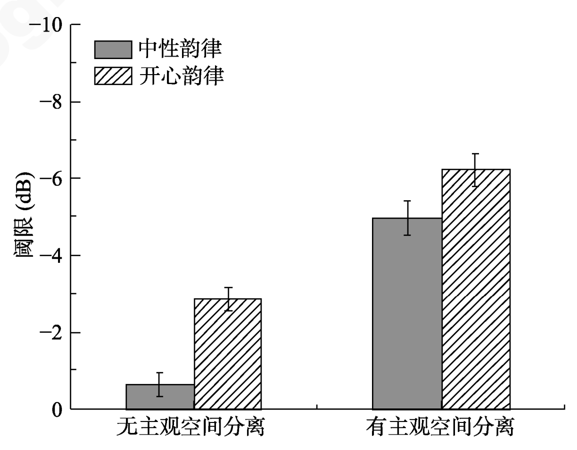


图 5　正序言语掩蔽下的识别阈限

单因素方差分析表明，听觉线索类型效应显著，F (2, 90) = 28.42, p < 0.001。情绪韵律的掩蔽释放量(M = 2.22, 95% CI: [1.60, 2.83])显著低于主观空间分离(M = 4.33, 95% CI: [3.68, 4.98], p < 0.001)或二者结合时的释放量(M = 5.59, 95% CI: [4.90, 6.28], p < 0.001)；主观空间分离线索的掩蔽释放量也低于二者结合的效应，p = 0.007。结果如图 6 所示。

综上，实验 1b 发现在正序播放的无意义语句带来的知觉、认知双重信息掩蔽下，情绪韵律线索也具有去掩蔽作用。情绪韵律线索的作用依然受到主观空间分离和信噪比因素的影响，但相比实验 1a，在较低信噪比条件下，即使存在主观空间分离，

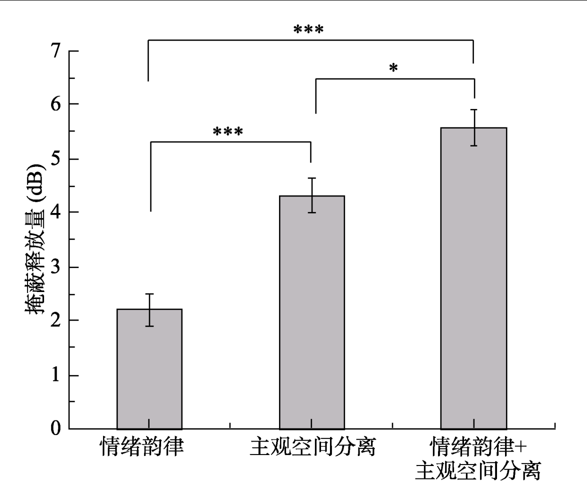


图 6　不同听觉线索的掩蔽释放量

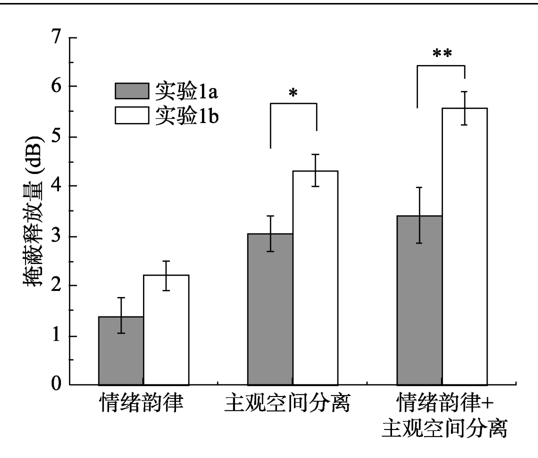


图 7　实验 1a 和实验 1b 去掩蔽量结果对比

情绪韵律线索依然具有去掩蔽作用。这可能是因为正序播放的无意义语句除了会导致知觉掩蔽外，还会有因语义可懂度所带来的认知掩蔽。更多的掩蔽量也就留下了更大的掩蔽释放空间，在信噪比较低的条件下，即使存在主观空间分离因素，目标声的情绪韵律依然可以帮助听者更好地加工目标声音，从而带来一定程度的掩蔽释放。然而，当听觉任务过于简单时，情绪韵律线索就不再有效了。

2.3　实验 1a 和实验 1b 结果的对比分析

为了探究在单纯知觉掩蔽下和在知觉、认知双重信息掩蔽下不同线索的去掩蔽效应是否存在差异，分别将实验 1a 中情绪韵律、主观空间分离和二者结合时的掩蔽释放量与实验 1b 所求得的相应掩蔽释放量进行独立样本 t 检验。结果表明，(1)对于情绪韵律线索，在知觉信息掩蔽下的掩蔽释放量和在知觉、认知双重掩蔽下的掩蔽释放量尽管存在差异，但没有达到 0.05 的显著性水平，t (55) = −1.75, p = 0.086; (2)对于主观空间分离线索，在知觉信息掩蔽下的掩蔽释放量显著小于在知觉、认知双重信息掩蔽下的掩蔽释放量，t (55) = −2.66, p = 0.010, 95% CI: [−2.23, −0.32]; (3)当情绪韵律和主观空间分离线索结合时，在知觉掩蔽下的掩蔽释放量和在知觉与认知信息掩蔽下的掩蔽释放量差异显著，t (55) = −3.41, p = 0.001, 95% CI: [−3.43, −0.89]。结果如图 7 所示。

这些结果表明单纯的情绪韵律线索主要对知觉信息掩蔽有效；主观空间分离因素既有去知觉掩蔽，也有去认知掩蔽的作用。有趣的是，情绪韵律线索和主观空间分离结合时也具有去认知信息掩蔽的作用，这可能是由于主观空间分离线索在起主要作用。

3　实验 2：情绪语义的去掩蔽作用

实验 2 考察目标声音中所包含关键词的情绪性语义在时间逆转(实验 2a)和正序(实验 2b)言语掩蔽条件下的去掩蔽作用。

3.1　实验 2a：逆序言语掩蔽条件下情绪语义的去信息掩蔽作用

3.1.1　被试

31 名听力正常的河北大学在校学生参加实验，其中男生 14 人，女生 17 人。平均年龄为 19.16 ± 1.59 岁。母语为汉语，右利手，并通过纯音听力测试(听力计，Conera, GN OTOMETRICS A/S)，具有正常且左右平衡的听力(左右耳相差不高于 15 dB，在任何一个频率上不高于 20 dB)。实验结束后，获得一定报酬。在数据分析阶段剔除一名数据异常的被试，有 30 名被试的数据进入最后的统计分析。

3.1.2　材料和设备

目标刺激为由一个特定说话人读出的汉语无意义语句，但与实验 1 相比在两个方面存在差异。首先，实验二中所有目标语句都为中性情绪韵律。其次，在实验二中的积极语义条件下，无意义语句的主语、谓语和宾语成分均具有积极情绪效价，如“那些**情侣**可能**爱慕**这些**别墅**”。中性语义条件则与实验 1 中相同。具体地，选择在实验 1a 中词汇效价评定时评分为 4.50~5.50，唤醒度相对较低的动词(107 个，评分 ≤ 3.20)和名词(214 个，评分 ≤ 2.69)作为中性词；评分为 5.80~9.00，唤醒度相对较高的动词(107 个，评分 ≥ 5.26)和名词(214 个，评分 ≥ 4.23)作为积极情绪词。对所有动词和名词的效价和唤醒度的评分进行分析。结果如表 3 所示。

分析表明，积极效价动词与中性效价动词在效价和唤醒度上差异显著{效价：t (287) = −42.00, p <

表 3 动词和名词的效价和唤醒度($M \pm SD$)

情绪语义	动词		名词	
	效价	唤醒度	效价	唤醒度
中性	5.07 ± 0.19	3.26 ± 0.62	5.03 ± 0.14	2.62 ± 0.33
积极	6.79 ± 0.40	5.99 ± 0.52	6.46 ± 0.32	5.06 ± 0.53

0.001, 95% CI: [−1.80, −1.64]; 唤醒度：t (287) = −38.38, $p < 0.001$, 95% CI: [−2.87, −2.59]}。积极效价名词与中性效价名词在效价和唤醒度上差异也均显著{效价：t (576) = −61.43, $p < 0.001$, 95% CI: [−1.48, −1.39]; 唤醒度：t (576) = −60.49, $p < 0.001$, 95% CI: [−2.51, −2.35]}。

为了确保实验 2 中不同情绪语义的句子在韵律上保持一致，由 23 名和 20 名未参与实验的被试分别对中性和积极语义条件下目标句的韵律效价和唤醒度进行评分，结果发现，在感知声音韵律的效价上，情绪语义效应不显著，t (212) = 1.53, p = 0.127, 95% CI: [−0.00, 0.03]; 在感知声音韵律的唤醒度上，情绪语义效应也不显著 t (212) = 0.46, p = 0.650, 95% CI: [−0.05, 0.08]。

掩蔽刺激为 2 个说话人说出的时间逆转的无意义语句(Yang et al., 2007)，与实验 1a 相同。实验设备同实验 1a。

3.1.3 设计和程序

实验采用 2×2×4 的被试内设计。3 个被试内变量及水平分别为：(1)主观空间分离(有主观空间分离、无主观空间分离); (2)情绪语义内容(中性效价、积极效价); (3)信噪比(4 dB、0 dB、−4 dB、−8 dB)。每个被试共接受 16 个实验条件，每个条件下有 11 个 trials。根据主观空间位置关系和语义条件，将所有 trials 分为 4 个 block, 4 个 block 的顺序在不同的被试间进行完全拉丁方平衡。4 种信噪比水平在每个 block 中随机化。因变量为被试复述关键词的正确率。

程序同实验 1a。

3.1.4 实验结果

使用 SPSS 21.0 软件对所有数据进行统计分析。

首先，对所有被试在不同条件下识别目标句的正确率进行均值分析，描述统计结果见表 4。

对不同实验条件下的识别正确率进行 3 因素的重复测量方差分析(所有 p 值都经过 Greenhouse-Geisser 矫正)。结果表明，主观空间分离主效应显著，F (1, 29) = 118.51, $p < 0.001$, $\eta_p^2 = 0.80$, 在主观空间分离条件下的识别正确率(M = 0.77, 95% CI:

表 4 识别目标句的正确率($M \pm SD$)

信噪比(dB)	无主观空间分离		有主观空间分离	
	中性语义	积极语义	中性语义	积极语义
−8	0.37 ± 0.12	0.35 ± 0.11	0.51 ± 0.16	0.52 ± 0.17
−4	0.58 ± 0.15	0.63 ± 0.12	0.73 ± 0.10	0.74 ± 0.11
0	0.83 ± 0.06	0.85 ± 0.09	0.87 ± 0.07	0.90 ± 0.06
4	0.93 ± 0.05	0.93 ± 0.05	0.94 ± 0.04	0.94 ± 0.05

[0.74, 0.80])高于在无主观空间分离条件下的识别正确率(M = 0.69, 95%, CI: [0.66, 0.71])。情绪语义主效应不显著，F (1, 29) = 2.60, p = 0.117。信噪比主效应显著，F (3, 87) = 509.62, $p < 0.001$, $\eta_p^2 = 0.95$; 随着信噪比的提高，识别准确率也提高，$ps < 0.001$。

主观空间分离和情绪语义交互作用不显著，F (1, 29) = 0.00, p = 0.989。主观空间分离和信噪比交互作用显著，F (3, 87) = 36.28, $p < 0.001$, $\eta_p^2 = 0.56$。简单效应分析发现，当信噪比为−8 dB、−4 dB 或者 0 dB 时，在有主观空间分离条件下的识别正确率显著高于无主观空间分离条件下的识别正确率(SNR 为−8 dB 或者−4 dB 时，$ps < 0.001$; SNR 为 0 dB 时，p = 0.001)。情绪语义和信噪比交互作用不显著，F (3, 87) = 1.53, p = 0.22。三者交互作用不显著，F (3, 87) = 1.13, p = 0.336。

同实验 1a, 将每个被试在不同实验条件下识别目标语句的正确率转化为相应的语音识别阈限。图 8 显示了所有被试平均识别率的拟合结果。不同主观空间分离和情绪语义条件下的平均识别阈限结果如图 9 所示。实验 2a 进而求得在知觉信息掩蔽下，情绪语义、主观空间分离线索和二者结合时的掩蔽释放量。

单因素方差分析表明，听觉线索类型效应显著，

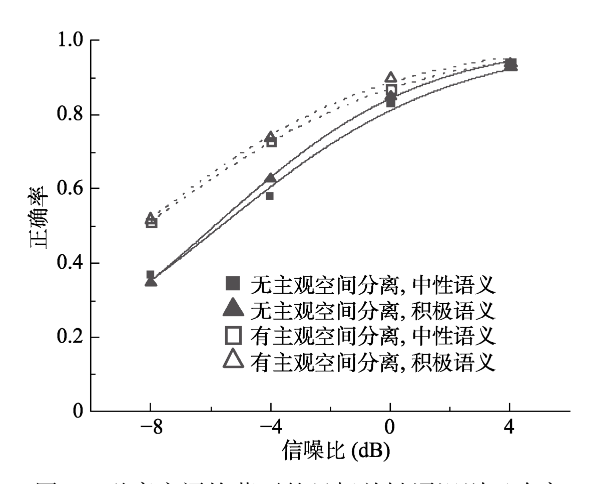


图 8 逆序言语掩蔽下的目标关键词识别正确率

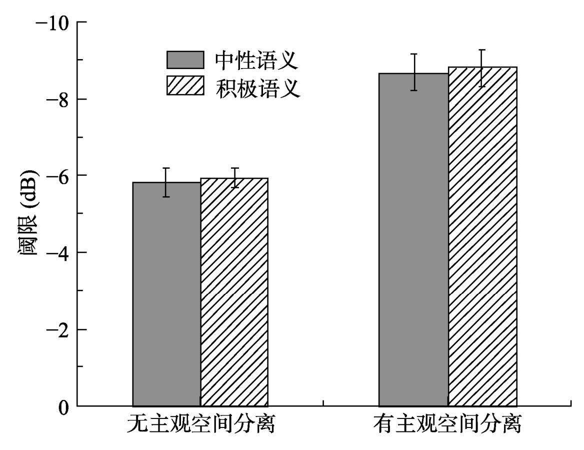


图 9　逆序言语掩蔽下的识别阈限

$F(2, 87) = 20.44$, $p < 0.001$。情绪语义的掩蔽释放量($M = 0.11$, 95% CI: [−0.67, 0.88])显著低于主观空间分离($M = 2.84$, 95% CI: [2.05, 3.63], $p < 0.001$)和二者结合的掩蔽释放量($M = 2.97$, 95% CI: [2.36, 3.58], $p < 0.001$), 但主观空间分离的掩蔽释放量和二者结合的效应差异不显著, $p = 0.796$。结果如图 10 所示。

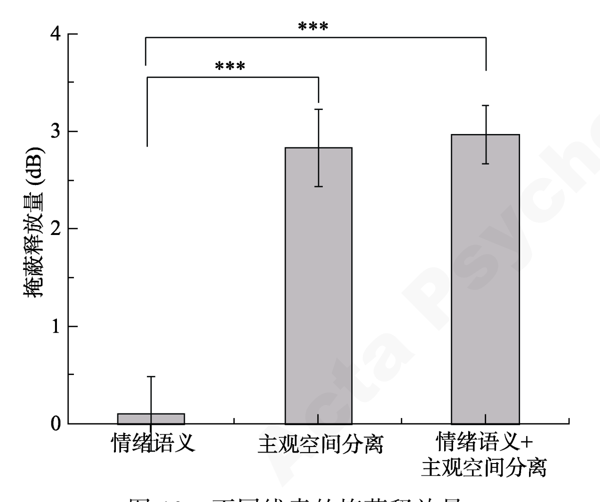


图 10　不同线索的掩蔽释放量

综上, 实验 2a 发现在知觉信息掩蔽下, 情绪语义线索本身不具有去掩蔽的作用, 并且与空间线索和听觉任务难度均没有交互作用。那在知觉、认知双重信息掩蔽下, 情绪语义是否具有去掩蔽的作用? 实验 2b 将对其进行探究。

3.2　实验 2b: 正序言语掩蔽条件下情绪语义的去信息掩蔽作用

3.2.1　被试

27 名听力正常的河北大学在校学生参加实验, 其中男生 10 人, 女生 17 人。平均年龄为 19.67 ± 1.75 岁。母语为汉语, 右利手, 并通过纯音听力测试(听力计, Conera, GN OTOMETRICS A/S), 具有正常且左右平衡的听力(左右耳相差不高于 15 dB, 在任何一个频率上不高于 20 dB)。实验结束后, 获得一定报酬。

3.2.2　材料和设备

目标刺激同实验 2a; 掩蔽刺激为正序播放的 2 个说话人说出的句法正确的汉语无意义语句(Yang et al., 2007)。实验设备同实验 1a。

3.2.3　设计和程序

均同实验 2a。

3.2.4　实验结果

使用 SPSS 21.0 软件对所有数据进行统计分析。

首先, 对所有被试在不同条件下识别目标句的正确率进行均值分析, 描述统计结果如表 5。

表 5　识别目标句的正确率($M \pm SD$)

信噪比(dB)	无主观空间分离		有主观空间分离	
	中性语义	积极语义	中性语义	积极语义
−8	0.06 ± 0.06	0.09 ± 0.06	0.27 ± 0.18	0.33 ± 0.18
−4	0.27 ± 0.11	0.32 ± 0.13	0.58 ± 0.15	0.64 ± 0.16
0	0.65 ± 0.10	0.74 ± 0.11	0.84 ± 0.08	0.88 ± 0.06
4	0.90 ± 0.05	0.93 ± 0.05	0.95 ± 0.04	0.95 ± 0.03

对不同实验条件下的识别正确率进行 3 因素的重复测量方差分析(所有 p 值都经过 Greenhouse-Geisser 矫正)。结果表明, 主观空间分离主效应显著, $F(1, 26) = 177.11$, $p < 0.001$, $\eta_p^2 = 0.87$, 在主观空间分离条件下的识别正确率($M = 0.68$, 95% CI: [0.65, 0.71])高于在无主观空间分离条件下的识别正确率($M = 0.50$, 95% CI: [0.48, 0.52])。情绪语义主效应显著, $F(1, 26) = 19.55$, $p < 0.001$, $\eta_p^2 = 0.43$, 在积极语义条件下的识别正确率($M = 0.61$, 95% CI: [0.58, 0.64])高于中性语义条件下的识别正确率($M = 0.57$, 95% CI: [0.54, 0.59])。信噪比主效应显著, $F(3, 78) = 1131.78$, $p < 0.001$, $\eta_p^2 = 0.98$; 随着信噪比的提高, 识别准确率也提高, $ps < 0.001$。

主观空间分离和情绪语义交互作用不显著, $F(1, 26) = 0.27$, $p = 0.611$。主观空间分离和信噪比交互作用显著, $F(3, 78) = 46.52$, $p < 0.001$, $\eta_p^2 = 0.641$。简单效应分析发现, 不管信噪比的大小, 无主观空间分离条件下的识别正确率均显著低于有主观空间分离条件下的识别正确率, $ps < 0.001$; 但不同信噪比条件下知觉空间分离所带来的效应量有较大差异, 信噪比从−8 dB 增加到 4 dB 条件时, Cohen's d 值分别为 1.29、1.99、1.35 和 0.54。情绪语义和信

噪比交互作用不显著, $F(3, 78) = 2.41$, $p = 0.077$。三因素交互作用不显著, $F(3, 78) = 1.88$, $p = 0.143$。

进一步将每个被试在不同实验条件下识别目标语句的正确率转化为相应的语音识别阈限。图 11 显示了所有被试平均识别率的拟合结果。不同主观空间分离和情绪语义条件下平均识别阈限结果如图 12 所示。

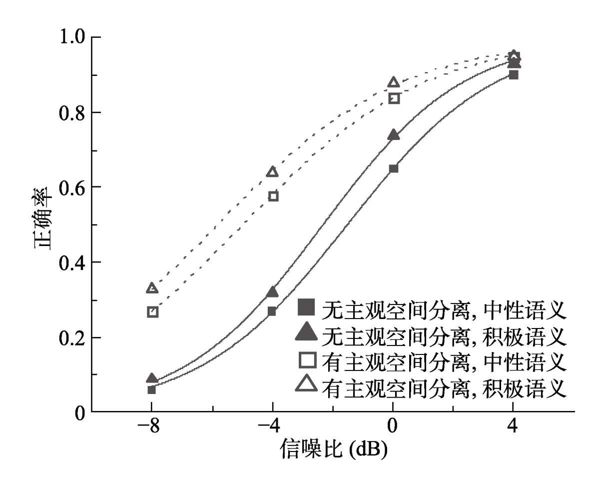


图 11 正序言语掩蔽下的目标关键词识别正确率

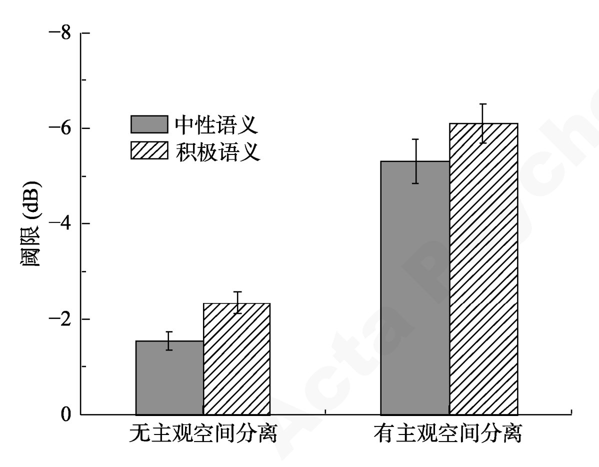


图 12 正序言语掩蔽下的识别阈限

进一步求得情绪语义、主观空间分离线索和二者结合时的掩蔽释放量。对其进行单因素方差分析(由于方差不齐性, 进行 Welch 方差分析)。结果表明不同线索条件下去掩蔽量差异显著, $F(2, 48.94) = 35.70$, $p < 0.001$, 积极语义去掩蔽量($M = 0.80$, 95% CI: [0.26, 1.35])低于主观空间分离($M = 3.75$, 95% CI: [2.84, 4.66], $p < 0.001$)和主观空间分离与积极语义的结合($M = 4.56$, 95% CI: [3.70, 5.41], $p < 0.001$); 主观空间分离和积极语义与主观空间分离的结合差异不显著, $p = 0.387$。结果如图 13 所示。

综上, 实验 2b 发现在知觉、认知双重信息掩蔽下, 情绪语义会起到较小但显著的去掩蔽作用。但同实验 2a 一样, 情绪语义线索不受空间线索和听觉任务难度的影响。

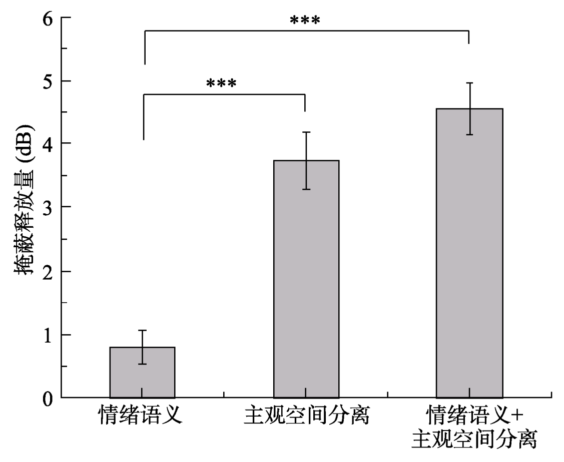


图 13 不同线索的掩蔽释放量

3.3 实验 2a 和实验 2b 结果对比分析

实验 2a 结果表明在知觉信息掩蔽下, 情绪语义不具有去掩蔽的作用, 而主观空间分离和主观空间分离与情绪语义结合时均具有去掩蔽的作用; 实验 2b 结果表明在知觉、认知双重信息掩蔽下, 情绪语义、主观空间分离和两者的结合均具有去掩蔽的作用。因此, 可以认为情绪语义不具有去知觉信息掩蔽的作用, 但具有去认知信息掩蔽的作用。之后分别将实验 2a 所求得的情绪语义、主观空间分离以及二者结合时的掩蔽释放量与实验 2b 所求得的情绪语义、主观空间分离和二者结合时的掩蔽释放量进行独立样本 t 检验。

结果表明, (1)情绪语义线索在知觉信息掩蔽下的掩蔽释放量与在知觉和认知信息掩蔽下的掩蔽释放量无显著差异, $t(55) = −1.48$, $p = 0.144$, 95% CI: [−1.64, 0.25]; (2)主观空间分离线索在知觉信息掩蔽下的掩蔽释放量与知觉和认知双重掩蔽下的掩蔽释放量差异不显著, $t(55) = −1.55$, $p = 0.13$, 95% CI: [−2.09, 0.27]; (3)当主观空间分离与积极语义线索结合时, 在知觉信息掩蔽下的掩蔽释放量和在知觉和认知双重掩蔽下的掩蔽释放量差异显著, $t(55) = −3.10$, $p = 0.003$, 95% CI: [−2.61, −0.56]。结果如图 14 所示。

4 总讨论

4.1 言语的情绪韵律去听觉信息掩蔽的机制

实验 1 和实验 2 分别考察了目标言语的情绪韵律与情绪语义信息在去信息掩蔽中的作用。实验 1a

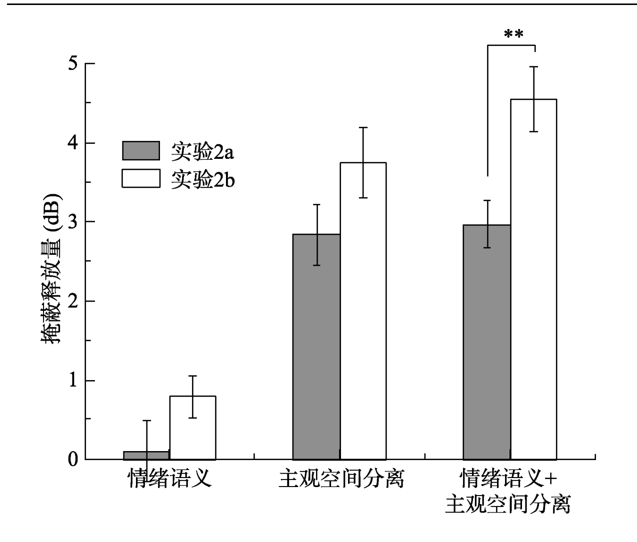


图 14　实验 2a 和实验 2b 去掩蔽量结果对比

和实验 1b 结果的对比分析表明，言语的情绪韵律具有去知觉信息掩蔽的作用，但对认知信息掩蔽作用很小。这能够进一步解释 Dupuis 和 Pichora-Fuller (2014)研究的结果。根据听觉注意理论，听者能够通过一些声学特点将复杂的听觉输入分解成独立的听觉客体，不同的听觉客体会竞争注意资源以主导听觉知觉。因为有情绪韵律的言语具有更为特殊的声学特征，例如，音高的平均值和范围较高(Dupuis & Pichora-Fuller, 2014)，能够为优先感知和注意的分配提供线索，所以能够在复杂的听觉场景中获得加工优势(Asutay & Västfjäll, 2014; Fritz et al., 2007; Shinn-Cunningham, 2008)，起到去知觉信息掩蔽的作用。这种去掩蔽的作用在很大程度上是独立于目标言语的内容的，可以认为是一种前认知过程。

吴超等人(2013)认为掩蔽环境下的言语感知依赖于两个过程。其中之一是客体的内在特征即信号的频谱−时间结构，它能够帮助个体对客体信号进行分流，以形成对客体的印象。因此，情绪韵律可能以其独特的频谱−时间结构，优先促进听者对情绪韵律印象的形成，从而更好地感知目标言语，起到去掩蔽的作用。

然而另一方面，正如引言部分所提到的，言语的情绪信息不一定非要以声学特征的变化为载体，经过消极情绪条件化的声音即便在声学特征上与经过中性情绪条件化的声音没有任何区别，依然可以起到一定的去掩蔽作用(Lu et al., 2018)。本研究 1 中通过操纵说话者的情绪韵律观察到的去掩蔽作用是否全部是声学因素所致，还是还包含了声学特征之外的其他效应(如动机因素)，本研究的结果并不能在这两种可能性间作出区分。后续的研究可以考虑这一问题。

本实验结果也能够解释 Brungart (2001)发现的目标声和掩蔽声说话人的声音特点影响对目标声音识别的现象。例如，与所有声音的性别相同时相比，当说话者性别不同时，识别正确率更高；与目标声和掩蔽声由同一人发出时相比，当声音由性别相同的不同说话者发出时，识别正确率更高。这种现象可能是由于当目标声和掩蔽声在声学特征上有差别时，目标声音能够优先吸引更多的注意，从而达到了去知觉信息掩蔽作用的效果。此外，Brungart 等人(2001)发现，在有 3 或 4 个声音源的听觉环境中，当目标掩蔽比(TMR，目标声和每一个掩蔽声的强度之比)接近 0 dB 时，与所有声音均由相同性别发出时产生的信息掩蔽量相比，一个掩蔽声和其它声音(目标声和另外的掩蔽声)由不同性别发出时产生的信息掩蔽量更多。这可能是因为当目标声和其它掩蔽声声学特征相同时，仅一个掩蔽声的声学特征不同，其容易优先吸引更多的注意，进而增加了知觉信息掩蔽。

4.2　言语的情绪语义去听觉信息掩蔽的机制

实验 2b 的结果首次证明了情绪语义具有去信息掩蔽的作用。而且，综合实验 2a 和实验 2b 的结果发现，情绪语义不具有去知觉信息掩蔽的作用，仅具有去认知信息掩蔽的作用。因为情绪反应系统组织是由欲望和防御两个基本的动机系统组成(Lang et al., 1997; Lang & Bradley, 2010)，对于一个积极刺激的快速反应可能将获得奖赏的可能性最大化。此外，从生物进化的角度，人们为了生存和快速适应环境会对于具有威胁的和有利于维持生命的环境进行快速侦探。因此，具有积极语义的目标声音会优先占据更多的认知加工资源，从而可以在有语义信息干扰的情况下得到一定程度的优先加工。

Grimm 等人(2012)的研究使用 2-back 任务发现，视觉呈现的具有情绪性的单词能够激活与工作记忆加工过程有关的外侧前额叶区(lateral prefrontal cortex, Hampson et al., 2010)。此外，Song 等人(2017)使用情绪 Stroop 任务发现，与中性刺激相比，具有情绪特征的刺激干扰与任务有关信息的加工过程，产生认知冲突时(Mathews, 1990; LeDoux, 2000)，会激活外侧前额叶区，尤其是背外侧前额叶(dorsolateal prefrontal cortex, DLPEC)和额下回(inferior frontal gyrus, IFG)。这些结果都意味着情

绪性刺激能够优先占据更多的认知资源。那么，与中性语义相比，具有情绪语义的声音在有干扰声的背景下，是否也能够激活外侧前额叶区，未来的研究可以使用 fMRI 技术对其进行探究。

实验 2a 结果表明在不具语义可懂度的时间逆转语句的干扰下，情绪语义不具有释放掩蔽的作用；实验 2b 结果表明在句法正确的无意义语句的干扰下，情绪语义具有释放掩蔽的作用。这与实验 1 显示出完全不同的结果模式。这很可能是由于实验 2b 中的掩蔽材料尽管在整体结构上是无意义语句，但是从词汇水平来看仍然是有意义的，因而会对目标语音造成额外的认知掩蔽。而因为目标声音中的关键词包含的情绪性信息，具有一定的语义突显性，因而能够获得一定程度的优先加工，从而表现出小幅的抗掩蔽性。

4.3 言语的情绪信息和主观空间分离结合时去听觉信息掩蔽的机制

主观空间分离是去听觉掩蔽的主要线索之一。本实验结果表明主观空间分离具有去信息掩蔽的作用，并且大于情绪线索的作用，这与 Lu 等人(2018)的研究结果一致。实验 1a 和实验 2a 结果均表明主观空间分离具有去知觉信息掩蔽的作用。

情绪信息和主观空间分离线索结合时既具有去知觉信息掩蔽的作用又具有去认知信息掩蔽的作用。正序和逆序的句子有很大的差别，不仅在语义上，在声学特征上也有一定的差别(Rhebergen et al., 2005)。Rhebergen 等人的研究发现，对于说荷兰语的听者，与正序的瑞典语的相比，掩蔽声为时间逆转的瑞典语时听者的识别阈限提高了 2.3 dB。这说明了在排除语义可懂度的影响下，与正序语句相比，时间逆转的语句会造成更大的知觉信息掩蔽。因此，与不具语义可懂度的时间逆转语句相比，在句法正确的无意义语句的干扰下，知觉信息的掩蔽量可能会减小。然而语义信息和知觉信息可能存在交互作用，这也许会增加知觉信息掩蔽的效果，但对于这个问题，目前还缺乏相关的研究能够直接回答。因此，情绪韵律和主观空间分离线索结合时，在不具语义可懂度的时间逆转语句的干扰下的掩蔽释放量与在句法正确的无意义语句的干扰下的掩蔽释放量差异显著，这更能说明两个线索结合时去认知信息掩蔽的作用。

在去知觉信息掩蔽量上，情绪韵律与主观空间分离线索结合时的作用和单独的主观空间分离线索的作用差异不显著。这可能是由于主观空间分离线索优先使听者注意目标声音，抵抗了掩蔽声在知觉上的干扰，导致目标声音的情绪韵律不起作用。同样地，空间线索能够使目标声音优先占据更多的加工资源，导致在知觉、认知双重掩蔽下情绪语义线索的掩蔽释放量和情绪语义与主观空间分离线索结合时的效应没有显著差别。

5 结论

本研究通过 2 个实验分别探究了言语的情绪韵律和情绪语义去信息掩蔽的具体机制。实验结果表明，言语的情绪韵律能够优先吸引听者更多的注意，主要对知觉信息掩蔽起作用。言语的情绪语义能够优先获取听者更多和内容加工相关的高级认知加工资源，具有去认知信息掩蔽的作用，而不具有去知觉信息掩蔽的作用。言语的情绪韵律和情绪语义去信息掩蔽的认知心理机制不同。

参 考 文 献

Arbogast, T. L., Mason, C. R., & Kidd Jr, G. (2002). The effect of spatial separation on informational and energetic masking of speech. *The Journal of the Acoustical Society of America, 112*(5), 2086–2098.

Asutay, E., & Västfjäll, D. (2014). Emotional bias in change deafness in multisource auditory environments. *Journal of Experimental Psychology: General, 143*(1), 27–32.

Bestelmeyer, P. E., Maurage, P., Rouger, J., Latinus, M., & Belin, P. (2014). Adaptation to vocal expressions reveals multistep perception of auditory emotion. *Journal of Neuroscience, 34*(24), 8098–8105.

Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *The Journal of the Acoustical Society of America, 109*(3), 1101–1109.

Brungart, D. S., Simpson, B. D., Ericson, M. A., & Scott, K. R. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *The Journal of the Acoustical Society of America, 110*(5), 2527–2538.

Carlile, S., & Corkhill, C. (2015). Selective spatial attention modulates bottom-up informational masking of speech. *Scientific Reports, 5*(1), 8662.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America, 25*(5), 975–979.

Cooke, M., Garcia Lecumberri, M. L., & Barker, J. (2008). The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception. *The Journal of the Acoustical Society of America, 123*(1), 414–427.

Culling, J. F., & Mansell, E. R. (2013). Speech intelligibility among modulated and spatially distributed noise sources. *The Journal of the Acoustical Society of America, 133*(4), 2254–2261.

Dupuis, K., & Pichora-Fuller, M. K. (2014). Intelligibility of emotional speech in younger and older adults. *Ear and Hearing, 35*(6), 695–707.

Ethofer, T., Bretscher, J., Gschwind, M., Kreifelts, B.,

Wildgruber, D., & Vuilleumier, P. (2012). Emotional voice areas: Anatomic location, functional properties, and structural connections revealed by combined fMRI/DTI. *Cerebral Cortex, 22*(1), 191–200.

Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2001). Spatial release from informational masking in speech recognition. *The Journal of the Acoustical Society of America, 109*(5), 2112–2122.

Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2004). Effect of number of masking talkers and auditory priming on informational masking in speech recognition. *The Journal of the Acoustical Society of America, 115*(5), 2246–2256.

Freyman, R. L., Helfer, K. S., McCall, D. D., & Clifton, R. K. (1999). The role of perceived spatial separation in the unmasking of speech. *The Journal of the Acoustical Society of America, 106*(6), 3578–3588.

Fritz, J. B., Elhilali, M., David, S. V., & Shamma, S. A. (2007). Auditory attention—focusing the searchlight on sound. *Current Opinion in Neurobiology, 17*(4), 437–455.

Frühholz, S., & Grandjean, D. (2013). Processing of emotional vocalizations in bilateral inferior frontal cortex. *Neuroscience & Biobehavioral Reviews, 37*(10), 2847–2855.

Glyde, H., Buchholz, J., Dillon, H., Best, V., Hickson, L., & Cameron, S. (2013). The effect of better-ear glimpsing on spatial release from masking. *The Journal of the Acoustical Society of America, 134*(4), 2937–2945.

Goh, W. D., Yap, M. J., Lau, M. C., Ng, M. M., & Tan, L. C. (2016). Semantic richness effects in spoken word recognition: A lexical decision and semantic categorization megastudy. *Frontiers in Psychology, 7*, 976.

Gordon, M. S., & Hibberts, M. (2011). Audiovisual speech from emotionally expressive and lateralized faces. *Quarterly Journal of Experimental Psychology, 64*(4), 730–750.

Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., & Vuilleumier, P. (2005). The voices of wrath: Brain responses to angry prosody in meaningless speech. *Nature Neuroscience, 8*(2), 145–146.

Grass, A., Bayer, M., & Schacht, A. (2016). Electrophysiological correlates of emotional content and volume level in spoken word processing. *Frontiers in Human Neuroscience, 10*, 326.

Grimm, S., Weigand, A., Kazzer, P., Jacobs, A. M., & Bajbouj, M. (2012). Neural mechanisms underlying the integration of emotion and working memory. *Neuroimage, 61*(4), 1188–1194.

Hampson, M., Driesen, N., Roth, J. K., Gore, J. C., & Constable, R. T. (2010). Functional connectivity between task-positive and task-negative brain areas and its relation to working memory performance. *Magnetic Resonance Imaging, 28*(8), 1051–1057.

Huang, Y., Huang, Q., Chen, X., Wu, X., & Li, L. (2009). Transient auditory storage of acoustic details is associated with release of speech from informational masking in reverberant conditions. *Journal of Experimental Psychology: Human Perception and Performance, 35*(5), 1618–1628.

Iwashiro, N., Yahata, N., Kawamuro, Y., Kasai, K., & Yamasue, H. (2013). Aberrant interference of auditory negative words on attention in patients with schizophrenia. *PloS One, 8*(12), e83201.

Jaspers-Fayer, F., Ertl, M., Leicht, G., Leupelt, A., & Mulert, C. (2012). Single-trial EEG–fMRI coupling of the emotional auditory early posterior negativity. *Neuroimage, 62*(3), 1807–1814.

Johansson, K. (1997). The role of facial approach signals in speechreading. *Scandinavian Journal of Psychology, 38*(4), 335–341.

Kidd Jr, G., Arbogast, T. L., Mason, C. R., & Gallun, F. J. (2005). The advantage of knowing where to listen. *The Journal of the Acoustical Society of America, 118*(6), 3804– 3815.

Kidd, G., & Colburn, H. S. (2017). Informational masking in speech recognition. In A. B. Coffin & J. Sisneros (Eds.), *Springer Handbook of Auditory Research: Vol. 60: The auditory system at the cocktail party (*pp. 75–109). Springer-Verlag.

Lang, P. J., & Bradley, M. M. (2010). Emotion and the motivational brain. *Biological Psychology, 84*(3), 437–450.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1997). Motivated attention: Affect, activation, and action. In P. J. Lang, R. F. Simons, M. Balaban, & R. Simons (Eds.), *Attention and orienting: Sensory and motivational processes* (pp. 97–136). Google Play.

LeDoux, J. E. (2000). Emotion circuits in the brain. *Annual Review of Neuroscience, 23*(1), 155–184.

Lu, L., Bao, X., Chen, J., Qu, T., Wu, X., & Li, L. (2018). Emotionally conditioning the target-speech voice enhances recognition of the target speech under “cocktail-party” listening conditions. *Attention, Perception, & Psychophysics, 80*(4), 871–883.

Mathews, A. (1990). Why worry? The cognitive function of anxiety. *Behaviour Research and Therapy, 28*(6), 455–468.

Mattys, S. L., Brooks, J., & Cooke, M. (2009). Recognizing speech under a processing load: Dissociating energetic from informational factors. *Cognitive Psychology, 59*(3), 203–243.

Mittermeier, V., Leicht, G., Karch, S., Hegerl, U., Möller, H. J., Pogarell, O., & Mulert, C. (2011). Attention to emotion: Auditory-evoked potentials in an emotional choice reaction task and personality traits as assessed by the NEO FFI. *European Archives of Psychiatry and Clinical Neuroscience, 261*(2), 111–120.

Mothes-Lasch, M., Becker, M. P., Miltner, W. H., & Straube, T. (2016). Neural basis of processing threatening voices in a crowded auditory world. *Social Cognitive and Affective Neuroscience, 11*(5), 821–828.

Rakerd, B., Aaronson, N. L., & Hartmann, W. M. (2006). Release from speech-on-speech masking by adding a delayed masker at a different location. *The Journal of the Acoustical Society of America, 119*(3), 1597–1605.

Rhebergen, K. S., Versfeld, N. J., & Dreschler, W. A. (2005). Release from informational masking by time reversal of native and non-native interfering speech. *The Journal of the Acoustical Society of America, 118*(3), 1274–1277.

Sander, D., Grandjean, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., &Vuilleumier, P. (2005). Emotion and attention interactions in social cognition: Brain regions involved in processing anger prosody. *Neuroimage, 28*(4), 848–858.

Scott, S. K., & McGettigan, C. (2013). The neural processing of masked speech. *Hearing Research, 303*, 58–66.

Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends in Cognitive Sciences, 12*(5), 182–186.

Song, S., Zilverstand, A., Song, H., Uquillas, F. D. O., Wang, Y., Xie, C., ... Zou, Z. (2017). The influence of emotional interference on cognitive control: A meta-analysis of neuroimaging studies using the emotional Stroop task. *Scientific Reports, 7*(1), 2088.

Summers, R. J., & Roberts, B. (2020). Informational masking of speech by acoustically similar intelligible and

unintelligible interferers. *The Journal of the Acoustical Society of America, 147*(2), 1113–1125.
Sun, C. C., Hendrix, P., Ma, J., & Baayen, R. H. (2018). Chinese lexical database (CLD): A large-scale lexical database for simplified Mandarin Chinese. *Behavior Research Methods, 50*(6), 2606–2629.
Watson, C. S. (2005). Some comments on informational masking. *Acta Acustica United with Acustica, 91*(3), 502– 512.
Wilson, R. H., Trivette, C. P., Williams, D. A., & Watts, K. L. (2012). The effects of energetic and informational masking on the Words-in-Noise Test (WIN). *Journal of the American Academy of Audiology, 23*(7), 522–533.
Wu, C., Cao, S., Wu, X., & Li, L. (2013). Temporally pre-presented lipreading cues release speech from informational masking. *The Journal of the Acoustical Society of America, 133*(4), EL281–EL285.
Wu, C., Wu, X. H., & Li, L. (2013). Speech Recognition in Schizophrenic under Masking Conditions. *Advances in Psychological Science, 21*(6), 958–964.
[吴超, 吴玺宏, 李量. (2013). 精神分裂症患者在听觉掩蔽环境下的言语识别. *心理科学进展, 21*(6), 958–964.]
Wu, C., Zheng, Y., Li, J., Zhang, B., Li, R., Wu, H., ... Li, L. (2017). Activation and functional connectivity of the left inferior temporal gyrus during visual speech priming in healthy listeners and listeners with schizophrenia. *Frontiers in Neuroscience, 11*, 107.
Wu, X., Chen, J., Yang, Z., Huang, Q., Wang, M., & Li, L. (2007). Effect of number of masking talkers on speech-on-speech masking in Chinese. In *Eighth Annual Conference of the International Speech Communication Association, Antwerp, Belgium*, 390–393.
Yang, Z., Chen, J., Huang, Q., Wu, X., Wu, Y., Schneider, B. A., & Li, L. (2007). The effect of voice cuing on releasing Chinese speech from informational masking. *Speech Communication, 49*(12), 892–904.
Yang, Z. G., Zhang, T. T., Song, Y. W., & Li, L. (2014). The Subcomponents of Informational Masking: Evidence from Behavioral and Neural Imaging Studies. *Advances in Psychological Science, 22*(3), 400–408.
[杨志刚, 张亭亭, 宋耀武, 李量. (2014). 听觉信息掩蔽的亚成分：基于行为和脑成像研究的证明. *心理科学进展, 22*(3), 400–408.]

Unmasking effects of speech emotional prosody and semantics on auditory informational masking

ZHENG Xi[1], ZHANG Tingting[1], LI Liang[2], FAN Ning[1], YANG Zhigang[1]
([1] *College of Education, Hebei University, Baoding* 071002 *China*)
([2] *Department of Psychology, Peking University, Beijing* 100871 *China*)

Abstract

In daily communication, a speaker's voice usually carries a particular emotion. Emotional information is transmitted in two ways: prosody and the semantic content of speech. Previous studies have found that emotional prosody has the effect of releasing auditory masking. The purpose of the present study is a), to test whether the emotional semantic content also has the effect of releasing speech from informational masking, and if so, b) to explore what is the difference between the role of emotional prosody and emotional content in releasing informational masking.

This study consisted of two experiments, each divided into two sub-experiments. A perceived spatial separation paradigm was adopted in all experiments to separate the effects of informational masking from that of energetic masking. Experiment 1 explored the mechanism of emotional prosody in the unmasking of informational masking. A complete within-subject design of 2 (perceived spatial separation: no, have) × 2 (emotional prosody: neutral, happy) × 4 (signal-to-noise ratio: −8 dB, −4 dB, 0 dB, 4 dB) was adopted in both sub-experiments. Experiment 1a employed time- reversed sentences with no semantic intelligibility as masking sounds (with presumed only perceptual informational masking). Experiment 1b used syntactically correct nonsense sentences as masking sounds (with both perceptual and cognitive informational masking). Experiment 2 also contained two sub-experiments; it aimed to examine the role of the emotional semantics of speech in releasing informational masking. A complete within-subject design of 2 (perceived spatial separation: no, have) × 2 (emotional semantics: neutral, positive) × 4 (signal-to-noise ratio: −8 dB, −4 dB, 0 dB, 4 dB) was adopted in both sub-experiments. Experiment 2a employed time- reversed sentences with no semantic intelligibility as masking sounds. Experiment 2b used syntactically correct nonsense sentences as masking sounds.

Experiment 1a showed that the accuracy of recognition of the target sentence uttered in emotional prosody was significantly higher than that of the target sentence uttered in neutral prosody. Experiment 1b showed that

the accuracy of recognition of the target sentence uttered in emotional prosody was significantly higher than that of the target sentence uttered in neutral prosody. There was a marginally significant difference between the results of Experiment 1a and Experiment 1b. Experiment 2a showed no significant difference in recognition accuracy between target sentences with emotional semantics and those with neutral semantics. Experiment 2b showed that the recognition accuracy of target sentences with emotional semantics was significantly higher than that of target sentences with neutral semantics. The study found no significant difference between Experiments 2a and 2b.

In conclusion, the results of the present study suggest that the mechanisms of emotional prosody and emotional semantics is different in releasing speech from informational masking. Emotional prosody of speech can preferentially attract more attention from listeners and reduce perceptual informational masking, but it only has a minor effect on releasing cognitive informational masking. The emotional semantics of speech can preferentially occupy more cognitive processing resources of listeners. Hence, it can reduce the cognitive informational masking; however, it fails to release the perceptual informational masking.

Key words perceptual informational masking, cognitive informational masking, emotional prosody, emotional semantics, perceived spatial separation